

APRIL/MAY 2024

**CECA63A/CECS63A — BIG DATA
ANALYTICS**

Time : Three hours

Maximum : 75 marks

PART A — (10 × 2 = 20 marks)

Answer ALL questions.

1. Define Big Data.
2. What is meant by 'Velocity'?
3. Expand RTAP.
4. Define Sampling.
5. List out the Characteristics of Big Data.
6. What is 'Hadoop'?
7. State the features of 'MapReduce'.
8. Give the significance of 'Task Execution'.
9. Why 'HiveQL' is used in Hadoop?
10. Write the applications of Big Data.

PART B — (5 × 5 = 25 marks)

Answer ALL the questions.

11. (a) Explain the Challenges of Conventional Systems.

Or

- (b) Write short notes on Analysis vs. Reporting.

12. (a) Discuss about Sampling Data in a Stream.

Or

- (b) How to count distinct elements in a Stream?

13. (a) Give the Importance of Big Data.

Or

- (b) Describe the Patterns for Big Data deployment.

14. (a) Illustrate the various JAVA interfaces to HDFS.

Or

- (b) How to develop a MapReduce application? Explain.

15. (a) List out the Data processing operators in Pig.

Or

- (b) Discuss on Querying data in Hive.

PART C — (3 × 10 = 30 marks)

Answer any THREE questions.

16. Explain about Analytic process and tools.

17. Describe the Stream Data Model and Architecture.

18. Discuss about Application development in Hadoop.

19. Write a detailed note on HDFS.

20. Explain the IBM InfoSphere BigInsights